#2

PATENT
Docket No. FR9-2000-0059 (245)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

11011 U.S. PTO
09/977665
10/15/01

In re Application of ROMERO

Application No. Examiner:

Filed: (Herewith) Group Art Unit:

For: METHOD AND SYSTEM FOR SEMANTIC SPEECH RECOGNITION

CLAIM OF FOREIGN PRIORITY

Box Patent Application
Commissioner for Patents
Washington, D.C. 20231

Sir:

Priority under the International Convention for the Protection of Industrial Property and under 35 U.S.C. §119 is hereby claimed for the above-identified patent application, based upon European Application No. 00480123.9 filed December 20, 2000, and a certified copy of this application is submitted herewith which perfects the Claim of Foreign Priority.

Respectfully submitted,

Date: 10/15/01

Kevin T. Cuenot

Gregory A. Nelson
Registration No. 30,577
Kevin T. Cuenot
Registration No. 46,283
Steven M. Greenberg
Registration No. 44,725
AKERMAN SENTERFITT
222 Lakeview Avenue
Post Office Box 3188
West Palm Beach, FL 33402-3188
Telephone: (561) 653-5000

Eur päisches Patentamt | European Patent Office | Office eur péen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

00480123.9

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE, 24/01/01
LA HAYE, LE

Europäisches Patentamt | European Patent Office | Office européen des brevets

Blatt 2 d r B scheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 00480123.9
Demande n°:

Anmeldetag:
Date of filing: 20/12/00
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
INTERNATIONAL BUSINESS MACHINES CORPORATION
Armonk, NY 10504
UNITED STATES OF AMERICA

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Conceptual speech recognition system

In Anspruch genommene Prioriät(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

Conceptual Speech Recognition System

Field of the Invention

The present invention relates to a speech recognition system and, more particularly, to a system and method for performing Natural Language Understanding functions by directly identifying the semantic information and the data from the other information involved in a spoken utterance.

Background of the Invention

Generally, the conventional speech recognition systems that perform Natural Language Understanding functions operate in two main sequential stages. In a first stage, a speech recognition unit translates a speech in a text which contains the exact transcription of a user's utterance. In a second stage,

a specific unit of Natural Language Understanding (NLU) reads this text with the sequence of words that have been recognized in the first stage and generates the information required to process the request.

The prior art Natural Language techniques are based on a two-stage process which operates at the word level and which compares the words of the uttered speech to words previously stored in a word vocabulary.

The inventor has found a new and more efficient way to operate a speech recognition system for Natural Language applications wherein the specific Natural Language Unit and the associated computer resources are no longer required. The new speech recognition system accepts natural language utterances as input and generates directly all the information required to process a user's request.

Summary of the Invention

Therefore, it is an object of the present invention to provide a system and method to operate a speech recognition system in Natural Language applications.

It is another object of the present invention to provide a system and method which allows easy building of applications using limited computer resources, like embedded systems or digital signal processing systems, and having improved time response.

It is yet another object of the present invention to provide a speech recognition system and method to be used in multilingual applications and for applications that should be translated to other languages.

The accomplishment of these and other related objects is achieved by a system and method which configure a speech recognition system in such a way that it accepts natural

language utterances as input and generates directly as output a compound of specific data and 'semantic identifiers' also called hereinafter 'concepts'. The data are marked with tags that permits to distinguish the different kind of data. The semantic identifiers or concepts are represented by concept codes. The tags and the concept codes are defined during a preparation/training phase of the system and may be chosen independently of the language used in the application.

In particular, it is convenient in multilingual applications to choose concept codes having a common part in a unique language and specific parts to represent the specific language. For example, a common code for representing the concept of querying could be 'QUERY' associated to a specific code 'EN' for English, 'SP' for Spanish, etc.. It is then very easy to operate a final function relevant of the application and to simultaneously select the appropriate set of answers.

The decoding concept process allows that the concepts and the data are decoded from the utterance of one word or a plurality of words. The concept codes and the data tags may be as simple and short or as complex and long as required by the application. Moreover, a unique concept code may be associated to various combinations of words.

In particular embodiments, dummy codes are assigned to some information of an utterance recognized as having secondary importance as regard to major concepts.

Finally, in a preferred embodiment, a computer implemented speech recognition method for performing Natural Language Understanding (NLU) functions, comprises the steps of:

discretizing a user's utterance into a plurality of speech basic units, the user's utterance being a sequence of words in the form of a query or a command;

matching the plurality of speech basic units against a plurality of combinations of items, each item being either a tagged data or a concept code; and

generating the most likely combination of items representative of the user's utterance.

Preferably the matching step initially includes a first step of matching the speech basic units against a vocabulary of items, the vocabulary being a collection of individual items defined during a preparatory/training phase of the system. As well, the combinations of items are selected valid combinations of items defined during the preparatory/training phase of the system.

The novel features believed to be characteristic of this invention are set forth in the appended claims. The invention itself, however, as well as these and other related objects and advantages thereof, will be best understood by reference to the following detailed description to be read in conjunction with the accompanying drawings.

Brief Description of the Drawings

Figure 1 is a block diagram of a speech recognition system according to the present invention.

Figure 2 is a block diagram of the elements involved in the generation of the conceptual pronunciation dictionary and the conceptual syntax module.

Figure 3 is a block diagram of the elements involved in the generation of the target function identification module.

Detailed Description of the Preferred Embodiment

The method of the present invention is the result of two inventor's observations: the first one relies on the fact that when a speech recognition system is used in a dictation application, the basic items used are the words contained in the user utterance. On the contrary, with the conceptual speech recognition system of the invention, the important issues are the concepts and the data involved in the utterance not the concrete words used to express them. The second observation relies on the fact that most of the speech recognition systems are capable to operate with virtually any language due to the fact that every feature characterizing a language (such as the pronunciation, the vocabulary, the syntax,...) is defined in specific data files. One of these files defines the correspondence between the spelling of each word and his pronunciation. The novel idea that the inventor has found from these two observations is to define a specific language covering the scope of the user's application where the pronunciation of each word is exactly the same as the natural language and the spellings of the words are codes representing concepts and tags representing data. From this language's representation it becomes very easy to select and operate an appropriate function to execute a command or answer a question requested by the user.

Before describing entirely the system of the present invention, the new language description is illustrated on three basic user's utterances :

1.A first utterance type may be in the form of a query such as "Please, give me the phone number of Pedro Romero". With the conceptual analyzis, the term "Please" will be identified as a dummy word. The expression "Give me the phone number of" will be treated as a semantic identifier and it will be recognized

by a concept code "QUERY" (or "QUERY-EN" in order to indicate the english language in a multilingual application). Finally the combination "Pedro Romero" will be analyzed as a data and it will be for example tagged such as: "Pedro_fn Romero_ln" (where the tag _fn means that Juan is a firstname and the tag _ln means that Rojas is a lastname).

2.A second utterance type may be in the form of a command such as: "Please, transfer me to him" . In this sentence, there is no data. The expression "Transfer me to him" is a semantic identifier that will be recognized by a concept code "DIAL" (or "DIAL-EN" for english application).

3.A third utterance type may be an isolated data such as: "Pedro Romero". This expression is interpreted as a command utterance where the system understands "I want to speak to "Pedro Romero". As it will be detailed hereinafter, in such case the speech recognition system tags the utterance as "Pedro_fn Romero_ln" and the presumed concept code ("DIAL" in this example) for the 'silent' semantic identifier "I want to speak to" is added by a Target Function Identification Module.

Now referring initially to FIG. 1, a block diagram of a conceptual speech recognition system 100 according to a preferred embodiment of the present invention is shown operatively coupled to an application-specific module referred to as TFIM (Target Function Identification Module) 120. The conceptual speech recognition system 100 includes an acoustic processor 102 and an acoustic model 104 that are operatively coupled to a fast acoustic match 108 and a detailed acoustic match 110. The fast acoustic match 108 and detailed acoustic match 110, which are operatively coupled to each other, are collectively referred to as a decoder 106. A conceptual pronunciation dictionary 112 and a conceptual syntax module 114 are each operatively coupled to both the fast acoustic match 108 and the detailed acoustic match 110. Depending on

the application, the conceptual syntax module 114 can be implemented either as a conceptual language model 116 or as a conceptual grammar 118.

It is to be appreciated that the present invention is usable with any speech recognition system using a conceptual language model or conceptual grammar technology and is not, in any way, limited to use with or dependent on any details or methodologies of any particular speech recognition arrangement. For instance, even generalized speech recognition systems such as the commercially available large vocabulary Via Voice system from IBM Corporation may be adapted to permit and/or perform conceptual speech recognition functions in accordance with the invention. In any case, it should be understood that the elements illustrated in FIG. 1 may be implemented in various forms of hardware, software, or combinations thereof. As such, the main recognition elements (e.g., acoustic model 104, fast acoustic match 108, detailed acoustic match 110, conceptual pronunciation dictionary 112 and conceptual syntax module 114) are implemented in software on one or more appropriately programmed general purpose digital computers. Each general purpose digital computer may contain, for example, a central processing unit (CPU) operatively coupled to associated system memory, such as RAM, ROM and a mass storage device, via a computer interface bus. Accordingly, the software modules performing the functions described herein may be stored in ROM or mass storage and then loaded into RAM and executed by the CPU. As a result, FIG. 1 may be considered to include a suitable and preferred processor architecture for practicing the invention which may be achieved by programming the one or more general purpose processors. Of course, special purpose processors may be employed to implement the invention. Given the teachings of the invention provided herein, one of ordinary skill in the related art will be able to contemplate these and similar implementations of the elements of the invention.

A brief explanation of the functionality of the components of the conceptual speech recognition system 100 will now be given. The acoustic processor 102 receives the speech (a spoken words sequence) uttered by a speaker. As it is well known, such acoustic processor generates wave forms, transduces the utterances into an electrical signal, converts the electrical signal into a digital signal representative of the uttered speech, samples the speech signal and partitiones the signal into overlapping frames so that each frame is discretely processed by the remainder of the system. The output signal of the acoustic processor 102 is a combination of feature vectors from the input utterance and labels (or fenemes) from the feature vectors. The labels are considered in a general sense to identify a phone, a phone being the basic unit of a utterance.

The speech recognition process is constrained by the acoustic model 104 which corresponds to the phones employed in the system 100, the conceptual pronunciation dictionary 112 and the conceptual syntax module 114.

The conceptual pronunciation dictionary defines the pronunciation of every concept code and every tagged data (also called the items), and is preferably a file containing a list of the items used wherein each item is followed by the phones associated to its pronunciation.

The conceptual syntax module 114 specifies the allowable combinations of items, and may be implemented as a conceptual language model 116 or as a conceptual grammar 118.

Generally, in Speech Recognition Systems, the collection of words that the system is able to recognize is contained in a file called the vocabulary. In the system of the invention, the Speech Recognition System does not recognize the uttered words but instead the "concepts" and "data" and thus, within this invention, the vocabulary is not as in the prior art

systems, a list of words but a list of "items" defining "concept codes" and "tagged data".

The output of acoustic processor 102 (a string of labels identifying a corresponding sound type) is input to decoder 106 including the fast acoustic match 108 and the detailed acoustic match 110. The object of the fast acoustic match 108 is to compare a string of incoming labels to the items stored in the conceptual vocabulary. The fast acoustic match initially recognizes items in the incoming labels and performs a reduction process to reduce the number of recognized items that require further processing. Preferably, the fast acoustic match is based on probabilistic finite state machines also well-known as the HMMs (Hidden Markov Models), and the candidate items are selected when acoustically similar to the stored items. A fast match candidate items list is thus produced from the fast acoustic match process.

Once the fast match reduces the number of candidate items, the fast match candidate items list is input to the detailed acoustic match module which determines in relation to the conceptual syntax module, the contextual likelihood of each candidate item based preferably on existing tri-grams. Preferably, the detailed acoustic match examines those items from the fast match candidate item list which have a reasonable likelihood of being the spoken item based on either the conceptual language model computation or the conceptual grammar. After the detailed match comparison, the conceptual syntax module is, preferably, again invoked to compute the likelihood of a segment of acoustics given the conceptual language model. The decoder of the present invention--using information derived from the fast matching, detailed matching, and applying the conceptual language model--is designed to determine the most likely path, or sequence, of items for a string of generated labels.

The output of decoder 106 is then a reduced list of decoded items resulting from both processes of the fast and the detailed acoustic match modules.

The decoded items output from the acoustic decoder 106 are provided to the application-specific Target Function Identification Module (TFIM) 120 which will execute the function corresponding to the decoded output. It is to be understood that the application-specific TFIM 120 may be any system that employs decoded speech signals as input. For example, the application-specific TFIM 120 may be a telephone modem system whereby the spoken utterances received by the conceptual speech recognition system 100 represent concepts and data to be electronically forwarded to a remote location. The recognized concepts and data could correspond to a command from a housewife and the remote location could be a home computer. Of course, the above application is merely an example and, as such, the present invention is not intended to be limited thereby.

A more detailed explanation of the functionality of some of the components of the conceptual speech recognition system 100 is now given. The acoustic model 104 is built and trained by analyzing speech samples of hundreds of speakers. The model contains a collection of acoustic prototypes. Each prototype corresponds to a gaussian distribution of a set of feature vectors associated with a phone. When a segment of a speech is input to the conceptual speech recognition system 100, the acoustic processor 102 examines the uttered speech in successive time intervals and a label is assigned to the interval based on a prototype of the acoustic model which is the closest. The closest prototype is determined by different measures of the feature vectors of the input segment speech. That is, based on the feature vector values generated during a centisecond interval for example, one acoustic prototype from

the set of acoustic prototypes included in the acoustic model is selected as being the closest.

In a preferred embodiment, the conceptual pronunciation dictionary 112 may be implemented as a table of items (concepts and data) corresponding to an application of interest (e.g., a conversational name dialer, a system for tourist information, a system for hotel information, etc.). Each item (concept or datum) in the dictionary vocabulary is represented by a sequence of phones which are combined to form the pronunciation of the item. This sequence of phones is generally referred to as the baseform of an item (concept or a datum).

The conceptual language model 116 which may be one implementation of the conceptual syntax module 114 is built and trained by analyzing a large conceptual corpus as it will be further described with reference to figure 2. The conceptual language model consists of a collection of conditional probabilities corresponding to the combination of items in the vocabulary. The function of the conceptual language model is to express rules or restrictions as to the way the items are to be combined to form sentences. Preferably, the conceptual language model is a n-gram model which makes the assumption that the a-priori probability of an item sequence can be decomposed into conditional probabilities of each item given the n items preceding it. In the context of n-gram language models, a trigram is a string of three consecutive items (denoted by i1 i2 i3). Similarly, a bigram is a string of two consecutive items, and a unigram is a single item. The conditional probability of the trigram model may be expressed as Prob(i3|i2|i1).

An alternative implementation of the conceptual syntax module 114 is a conceptual grammar 118 which is designed to accept each valid combination of concepts and/or data contained in a conceptual corpus which will be detailed with reference to figure 2.

Before decoder 106 is used in real application and performs the utterance decoding process by using the feature vector signals and labels provided by acoustic processor 102, the acoustic model 104 and the conceptual language model 116 need to be trained. The parameters (probabilities) of both these models are generally estimated from training data from the application of interest. In order to train the acoustic model 104, acoustic training data are provided by a user of the system, as well as a transcription representative of the training data. The voices of many people are recorded. They speak in an environment similar to the environment where the system will be used and they record sentences similar to the sentences that the system will have to recognize. These sentences are transcript into text in order to make possible the association between the words used and how these words have been uttered. A statistical process extract the required information. The transcription may be input directly to decoder 106 as a text.

Further, in order to train the conceptual language model, a collection of sentences typical of an application domain is composed, and transcript into a text provided to the decoder. These sentences must be composed of valid sequences of items (concepts and data).

Preferably, a trigram language model as is well known by the skilled man is trained using a transcription consisting of a large corpus. The corpus consists of sentences. The training involves inputting the sentences and determining statistics for each item (concept or datum) model in a manner which enhances the probability of the correct item relative to the probabilities associated with other items. Such training provides counts for all trigrams, bigrams and unigrams identified in the conceptual corpus.

Figure 2 is a block diagram of the elements involved in the generation of the conceptual pronunciation dictionary 112 and

the conceptual syntax module 114. Three units (200,202,204) are used to define a specific application:

- a concept/word table 200,
- a word corpus 202, and
- a word pronunciation dictionary 204.

The concept/word table 200 contains every concept defined for every possible combination of words. The word corpus 202 contains real sentences that should be recognized, and the word pronunciation dictionary 204 contains sequences of phones reflecting the pronunciations of every word contained in the word corpus.

A conceptual corpus 206 is generated from the combination of the concepts/words contained in the concept/word table 200 with the words contained in the word corpus 202, by performing every possible translation defined in the concept/word table 200. And as already mentioned, the conceptual syntax module which uses the conceptual corpus may be in the form of a conceptual language model 116 or a conceptual grammar 118.

The conceptual pronunciation dictionary 112 is generated from the combination of the concept/word table 200 with the word pronunciation dictionary 204. The conceptual pronunciation dictionary 112 is obtained by replacing every word of the concept word table 200 by its corresponding pronunciation stored in the word pronunciation dictionary 204.

The skilled man will understand and thereby adapt the system to the case where the pronunciation of a concatenation of words does not correspond to the concatenation of the different pronunciations of the words.

Referring to FIG. 3, a block diagram of the elements involved in the generation of the target function identification module (TFIM) 120 is described.

A function/concept table 300 is built to store the functions to be executed in relation to every possible combination of concepts. The target function identification module (TFIM) 120 is an algorithm unit which performs the actions defined in the function/concept table 30. TFIM checks the decoder 106 output under specific conditions. A condition is a combination of concept codes. For example, if in a decoded sentence the concept codes "QUERY" and "PHONE" are identified, the TFIM module will execute the function "QUERY-PHONE-FUNCTION" passing the datum "NAME" as an argument in the call.

When a specific condition is identified, the appropriate function is called by TFIM from the function/concept table 300 and runs using the tagged data output by decoder 106. The TFIM module knows every relevant concept and data involved in the sentence, for that reason it can infer the global meaning using simple rules that can be implemented easily at a high level programming language, especially when the language provides built-in pattern matching and strings functions. Additionally, this module can perform a verification of the integrity and validity of the concepts and data recognized, so it can reject incompatible or uncertain combinations, thereby improving the application efficiency.

The method of the present invention requires that: a) every uttered sentence expresses at least a concept and data, or any sequential combination of both, b) every concept or data must be uttered as a continuous string of words and c) it should be possible to foresee all the alternative phrases which could be used to express every relevant concept and data.

As these conditions are slightly restrictive, the proposed method is of a general application. Besides, the method may be used for speech recognition software based on grammars as well

as language models. In any case the grammars or the corpus for training the language model must be defined using the selected codes for concepts and data. In the same way, this method may be used in any application where the voice is picked up by a microphone as well as by a phone.

The method of the present invention has been implemented to demonstrate its validity. A prototype was developed based on a telephony application already developed according to the traditional approach. The following paragraphs contain a brief description of both the traditional application (A) and the prototype application using the method of the present invention (B).

A) The original application was started from a telephony application in Spanish called Conversational Name dialer which was developed by IBM Corporation. Corresponding applications in German, French and English where also developed using the traditional approach as described below.

The Spanish version was installed on an IBM Personal Computer 300 PL equipped with a telephony adapter Dialogic D/41ESC. The application knows phone numbers of some 4000 Spanish IBM Corp. employees and was designed to answer calls from users asking for the phone number of an employee or requesting to transfer the call to one of them.

One of the main features of this application is that users do not have to know any special knowledge about how to use the application because it is designed, in this speaking environment, to understand and to answer sentences in a similar way that people do. For example, the user could address the application using formal or informal expressions, such as beginning a sentence with a greeting like "hello" / "good morning" / "good afternoon", etc. or he optionally can say who he is ("I'm Antonio García from IBM Madrid") or he could address the application using very short sentences or using different polite expressions.

The application was able to answer to user's questions and to establish dialogues when the request had some ambiguity. The application was using a text-to-speech (TTS) module to synthesize the voice. For example, if the user asked "Give me the phone number of Fernández" the application answered something like "There are many people so called, let me know some additional information". Then the user could answer with the first name and/or the location where he works and/or if he refers to a male or female employee, etc. This feature was implemented through a Dialog Manager (DM) module that provides the appropriate target function. The Speech Recognition task was using a language model based on words. The model was obtained from a corpus having sentences the users could be used. The operations to prepare the corpus were:

- to collect some 400 different sentences used within this environment,

- to select the elementary phrases (fragments) contained in those sentences,

- to generate new sentences obtained by mixing the fragments in valid combinations,

- to select words contained in the phone database (about 7000 first names, last names, cities and countries), and

- to generate new sentences from the previous list using the words selected from the phone database.

B) The prototype application using the method of the present invention was developed from the original application as described above. This was performed by replacing the speech recognition system by a conceptual speech recognition system that directly recognizes concepts and data from the user's utterance. In order to prepare the new conceptual speech recognition configuration, an analysis of the original word corpus had been performed. It has lead to the generation of 58

classes of items (concepts codes and tags) which allow to represent every sentence in the corpus in an alternative way. Every item is defined by a set of words and/or phrases having a similar meaning or a similar role in the sentences. A conceptual corpus was generated modifying the sentences by replacing every word or phrase by the corresponding concept code and replacing every data by the corresponding tagged data.

Within those 58 classes, 44 were representing concepts codes and 14 were representing tags.

The following table is an extract of some of the 44 concept codes of the prototype application and alternative sentences:

concepts codes	alternative expressions
HELLO	hello - good morning - good afternoon - good night - hello good morning - hello good afternoon - hello good night - ....
POLITE	please
QUESTION	what's - I'm calling to ask - I want to know - I want to confirm - I'd like to know - I'd like to confirm - ...
PHONE	the number of - the phone number of - the phone of -the extension number of - ...
LOCATION	from - he lives in - she lives in - he works in - she works in - ...

The following table is an extract of some of the 14 tags for the prototype application :

data	tag	tagged data
Pedro Romero	FIRSTNAME LASTNAME	Pedro(FIRSTNAME)- Romero(LASTNAME)- Pedro(FIRSTNAME)Romero(LASTNAME)
Maria Fernandez	FIRSTNAME LASTNAME	María(FIRSTNAME)- Fernández(LASTNAME)- María(FIRSTNAME)Fernández(LASTNAME)
Madrid	CITY	Madrid(CITY)

The following list shows some of the sentences generated for the prototype conceptual corpus:

HELLO QUESTION PHONE Pedro(FIRSTNAME) POLITE

HELLO QUESTION PHONE María(FIRSTNAME)Fernández(LASTNAME) POLITE

QUESTION PHONE Pedro(FIRSTNAME)Romero(LASTNAME) POLITE

PHONE Pedro(FIRSTNAME)Romero(LASTNAME)

PHONE DUMMY María(FIRSTNAME)Fernández(LASTNAME)

HELLO QUESTION PHONE Pedro(FIRSTNAME) LOCATION Madrid(CITY) POLITE

QUESTION DUMMY PHONE Pedro(FIRSTNAME)Romero(LASTNAME) DUMMY LOCATION Sevilla(CITY)

PHONE María(FIRSTNAME)Fernández(LASTNAME) DUMMY LOCATION Sevilla(CITY).

Claims

1. A computer implemented speech recognition method for performing Natural Language Understanding (NLU) functions, comprising the steps of:

discretizing a user's utterance into a plurality of speech basic units, the user's utterance being a sequence of words to express a query or a command;

matching the plurality of speech basic units against a plurality of combinations of items, each item being either a tagged data or a concept code; and

generating the most likely combination of items representative of the user's utterance.

2. The method of claim 1 wherein the matching step comprising a first step of matching the plurality of speech basic units against a vocabulary of items to generate a first list of most likely items representative of the user's utterance.

3. The method of claim 2 wherein the first matching step is performed using Hidden Markov Models.

4. The method of claims 2 or 3 wherein the matching step further comprising a second step of matching the first list of most likely items against the plurality of combinations of items to generate the most likely combination of items representative of the user's utterance.

5. The method of claim 4 wherein the second matching step is processed using a conceptual language model.

6. The method of claim 5 wherein said conceptual language model is a n-gram conceptual language model.

7. The method of claim 6 or 7 further comprising an initial step of training the conceptual language model.

8. The method of claim 4 wherein the second matching step is processed using a conceptual grammar.

9. The method of anyone of claims 2 to 8 further comprising a training step to define the vocabulary of items.

10. The method of claims 1 or 9 further comprising a training step to define the plurality of combinations of items.

11. The method of anyone of claims 1 to 10 further comprising a step of storing a set of prototypes acoustic models obtained from a training phase, each acoustic model representing one or more possible basic speech units of an utterance of a word.

12. The method of claim 11 further comprising a step of assigning an acoustic model to each speech basic unit.

13. The method of anyone of claims 1 to 12 wherein said user's utterance being in the form of an isolated data.

14. The method of anyone of claims 1 to 13 wherein said tagged data includes two consecutive words.

15. The method of anyone of claims 1 to 14 further comprising a step of sending said most likely combination of items to a function identification module to perform the user's query or command.

16. A system comprising means adapted for carrying out the steps of the method according to anyone of the preceding claims 1 to 15.

17. In a speech recognition system designed for performing a user's command or for answering a user's query, a method for for performing Natural Language Understanding (NLU) functions, comprising the steps of anyone of claims 1 to 15.

18. A computer program product stored on computer usable medium, comprising computer readable program means for causing a computer to perform a method according to anyone of the preceding claims 1 to 15.

Conceptual Speech Recognition System

Abstract

The present invention discloses a computer implemented method to understand queries or commands spoken by users when they use natural language utterances similar to those that people use spontaneously to communicate. More precisely, the invention discloses a method that identifies the user's queries or commands from the general information involved in spoken utterances directly by the speech recognition system and not by a post-process as usual. In a phase of preparation of the system, a vocabulary of items representing data and semantic identifiers is created as well as a syntax module having valid combinations of items. When the system is in use, a user's utterance is first discretized into a plurality of speech basic units which are compared to the items in the vocabulary and a combination of items is selected according to the evaluation from the syntax module in order to generate the most likely sequence of items representative of the user's utterance. Finally the semantic identifiers and the data extracted from the user's utterance are used to call the appropriate function that process the user's request.

Figure 1.

THIS PAGE BLANK (USPTO)

FR9 2000 0059
ROJAS ROMERO
1/3

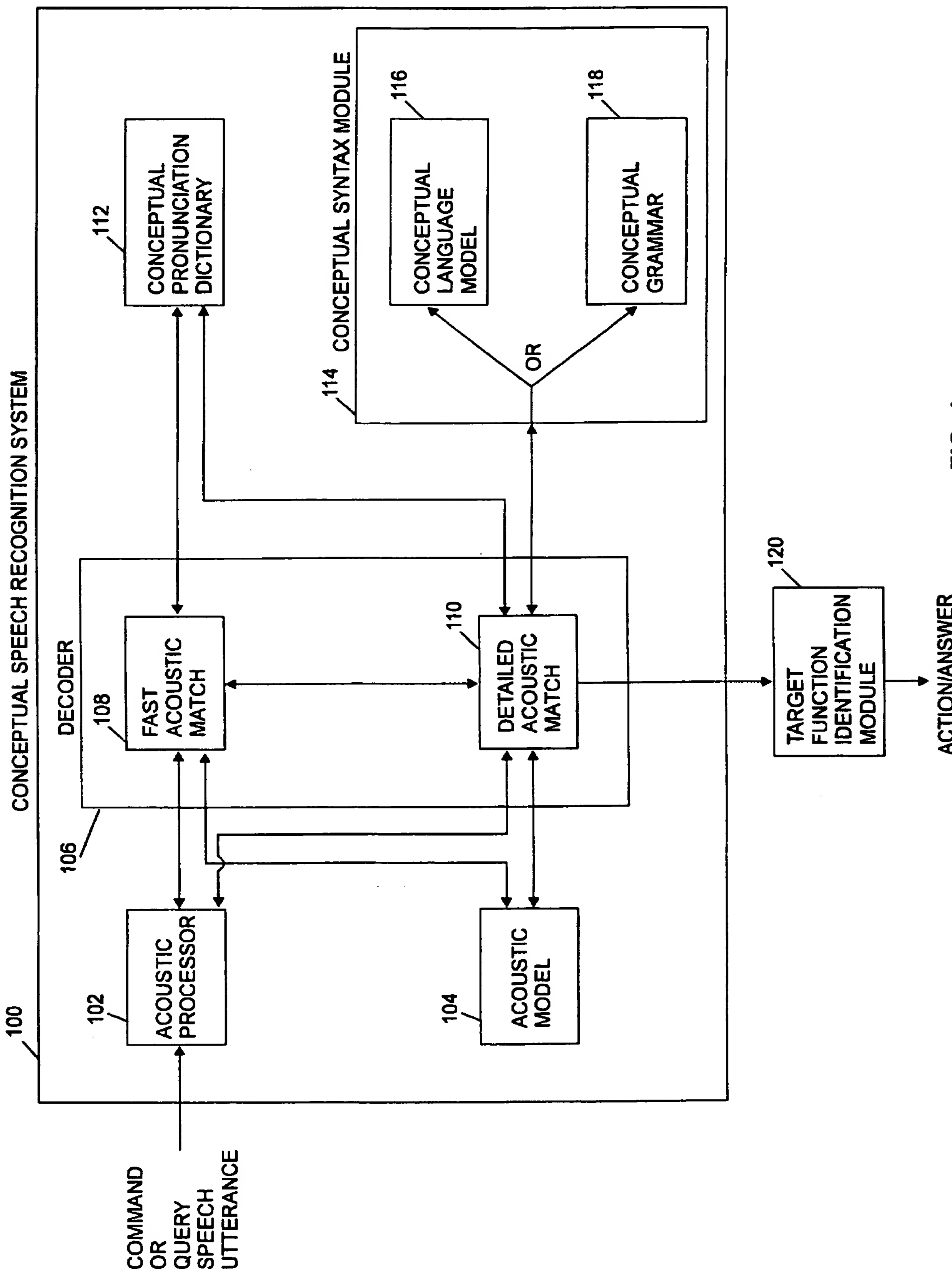

FIG. 1

FR9 2000 0059
ROJAS ROMERO
2/3

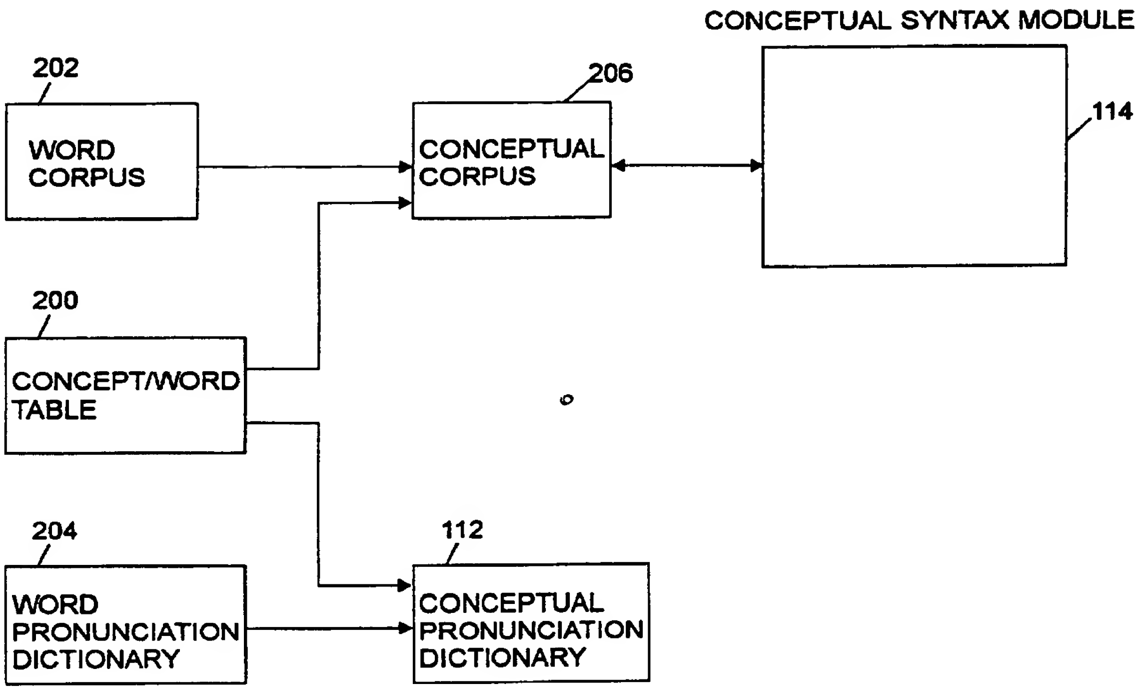

FIG. 2

FR9 2000 0059
ROJAS ROMERO
3/3

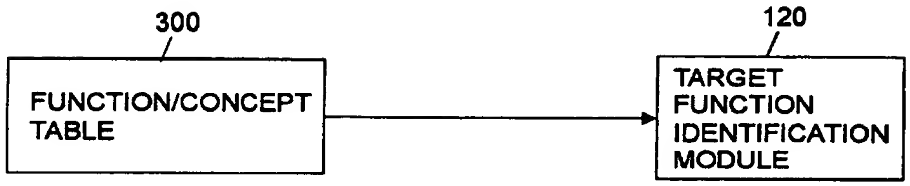

FIG. 3